Information and Computation 243 (2015) 249–262

Contents lists available at ScienceDirect

Information and Computation

www.elsevier.com/locate/yinco

A refined complexity analysis of degree anonymization in graphs ☆

Sepp Hartung [a,*], André Nichterlein [a,*], Rolf Niedermeier [a,*], Ondřej Suchý [b,*]

[a] *Institut für Softwaretechnik und Theoretische Informatik, TU Berlin, Germany*
[b] *Faculty of Information Technology, Czech Technical University in Prague, Czech Republic*

A R T I C L E I N F O

Article history:
Received 16 September 2013
Available online 13 December 2014

Keywords:
Parameterized complexity
Kernelization
Heuristics
f-Factors
Data privacy

A B S T R A C T

Motivated by a strongly growing interest in graph anonymization, we study the NP-hard Degree Anonymity problem asking whether a graph can be made k-anonymous by adding at most a given number of edges. Herein, a graph is k-anonymous if for every vertex in the graph there are at least $k-1$ other vertices of the same degree. Our algorithmic results shed light on the performance quality of a popular heuristic due to Liu and Terzi [ACM SIGMOD 2008]; in particular, we show that the heuristic provides optimal solutions if "many" edges need to be added. Based on this, we develop a polynomial-time data reduction yielding a polynomial-size problem kernel for Degree Anonymity parameterized by the maximum vertex degree. In terms of parameterized complexity analysis, this result is in a sense tight since we also show that the problem is already NP-hard for H-index three, implying NP-hardness for smaller parameters such as average degree and degeneracy.

1. Introduction

For many scientific disciplines, including the understanding of the spread of diseases in a globalized world or power consumption habits with impact on fighting global warming, the availability of social network data becomes more and more important. To respect privacy issues, there is a strong demand to anonymize the associated data in a preprocessing phase [21]. In a landmark paper, Liu and Terzi [31] (also see [14] for an extended version) introduced the following simple graph-theoretic model for identity anonymization on (social) networks. Herein, they transferred the k-anonymity concept known for tabular data in databases [21,41–43] to graphs (see Fig. 1 for examples).

Degree Anonymity [31]
Input: An undirected graph $G=(V,E)$ and two positive integers k and s.
Question: Is there an edge set E' over V with $|E'| \le s$ such that $G'=(V, E \cup E')$ is *k-anonymous*, that is, for every vertex $v \in V$ there are at least $k-1$ other vertices in G' having the same degree?

Liu and Terzi [31] assume in this model that an adversary (who wants to de-anonymize the network) knows only the degree of the vertex of a target individual; this is a modest adversarial model. Clearly, there are stronger adversarial models

☆ A preliminary version appeared in the *Proceedings of the 40th International Colloquium on Automata, Languages, and Programming (ICALP 2013), Part II*, Riga, Latvia, July 2013, Volume 7966 of Lecture Notes in Computer Science, pages 594–606, Springer.

* Corresponding authors.
E-mail addresses: sepp.hartung@tu-berlin.de (S. Hartung), andre.nichterlein@tu-berlin.de (A. Nichterlein), rolf.niedermeier@tu-berlin.de (R. Niedermeier), ondrej.suchy@fit.cvut.cz (O. Suchý).

http://dx.doi.org/10.1016/j.ic.2014.12.017
0890-5401/

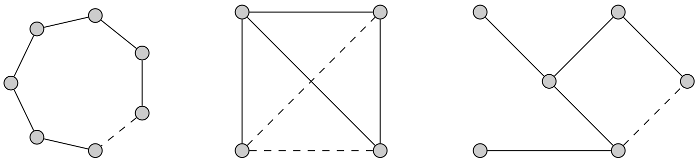

Fig. 1. Three illustrating examples. The solid edges indicate the original graphs. Adding the dashed edges changes the graphs (from left to right) from being 2-anonymous to 7-anonymous, from 1-anonymous to 4-anonymous, and from 1-anonymous to 2-anonymous.

which (in many cases very realistically) assume that the adversary has more knowledge, making it possible to breach privacy provided by a "k-anonymized graph" [36]. Moreover, it has been argued that graph anonymization has fundamental theoretical barriers which prevent a fully effective solution [1]. Degree Anonymity, however, provides the perhaps most basic and still practically relevant model for graph anonymization; it is the subject of active research [10,11,13,33].

Graph anonymization problems are typically NP-hard. Thus, almost all algorithms proposed in this field are heuristic in nature, this also being true for algorithms for Degree Anonymity [25,31,33]. Indeed, as the field of graph anonymization is young and under strong development, there is very little research on its theoretical foundations, particularly concerning computational complexity and algorithms with provable performance guarantees [11].

Our contributions. Our central result is to show that Degree Anonymity has a polynomial-size problem kernel when parameterized by the maximum vertex degree Δ of the input graph. In other words, we prove that there is a polynomial-time algorithm that transforms any input instance of Degree Anonymity into an equivalent instance with at most $O(\Delta^7)$ vertices. Indeed, we encounter a "win–win" situation when proving this result: We show that Liu and Terzi's heuristic strategy [31] finds an optimal solution when the size s of a minimum solution is larger than $2\Delta^4$. As a consequence, we can bound s in $O(\Delta^4)$ and, hence, a polynomial kernel we provide for the combined parameter (Δ, s) actually is also a polynomial kernel only for Δ. Furthermore, our kernelization has the useful property (for instance when combining it with approximation algorithms) that each solution derived for the kernel instance one-to-one corresponds to a solution of the original instance. While this kernelization directly implies fixed-parameter tractability for Degree Anonymity parameterized by Δ, we also develop a further fixed-parameter algorithm with an improved worst-case running time.

In addition, we prove that Degree Anonymity becomes NP-hard on graphs with H-index[1] three. The same proof also yields NP-hardness in 3-colorable graphs. Further, adopting the viewpoint of "standard parameterization", we show that Degree Anonymity is W[1]-hard when parameterized by the solution size s (the number of added edges), even when $k = 2$. In other words, there is no hope for fixed-parameter tractability even when the level k of anonymity is low and the graph needs only few edge additions (meaning little perturbation) to achieve k-anonymity.

Why is the parameter "maximum vertex degree Δ" of specific interest? First, note that from a parameterized complexity perspective it seems to be a "tight" parameterization in the sense that for the only little "stronger" (that is, provably smaller) parameter H-index our results already show NP-hardness for H-index three (also implying hardness e.g. for the parameters degeneracy and average degree). Social networks typically have few vertices with high degree and many vertices of small degree. Leskovec and Horvitz [30] studied a huge instant-messaging network (180 million vertices) with maximum degree 600. For the DBLP co-author graph[2] generated in February 2012 and containing more than 715,000 vertices we measured a maximum degree of 804 and an H-index of 208, so there are not more than 208 vertices with degree larger than 208. Thus, a plausible strategy might be to only anonymize vertices of "small" degree and to remove high-degree vertices for the anonymization process because it might be overly expensive to anonymize these high-degree vertices and since they might be well-known (that is, not anonymous) anyway. Indeed, high-degree vertices can be interpreted as outliers [2], potentially making their removal plausible.

Related work. The most important reference is Liu and Terzi's work [31] where the basic model of graph anonymization was introduced, sophisticated (heuristic) algorithms (also using algorithms to determine the realizability of degree sequences) have been developed and validated on experimental data. Somewhat more general models have been considered by Zhou and Pei [44] (studying the neighborhood of vertices instead of only the degree) and by Chester et al. [11] (anonymizing a *subset* of the vertices of the input). Chester et al. [13] investigated the variant of adding vertices instead of edges; Bredereck et al. [8] provided first parameterized complexity results in this direction. Recently, building on Liu and Terzi's work, we enhanced their heuristic approach with the focus on improving lower and upper bounds on the solution size [25]. Lu et al. [33] and Casas-Roma et al. [10] proposed enhanced algorithms for Degree Anonymity. Again, these algorithms are heuristic in nature. Today, the field of graph anonymization has grown tremendously with numerous surveys and research directions. We only mention some directly related work.

[1] The H-index of a graph G is the maximum integer h such that G has at least h vertices with degree at least h. As a consequence, if G has H-index h, then it has at most h vertices of degree larger than h.

[2] In this graph the vertices represent the authors and an edge indicates that the two corresponding authors are co-authors of at least one paper.